PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	How representative are colorectal, lung, breast and prostate cancer patients responding to the National Cancer Patient Experience Survey (CPES) of the cancer registry population in England? A population-based case control study
AUTHORS	Alessy, Saleh; Davies, Elizabeth; Rawlinson, Janette; Baker, Matthew; luchtenborg, Margreet

VERSION 1 - REVIEW

REVIEWER	Anna Gigli Institute for Research on Population and Social Policies, National Research Council, Italy
REVIEW RETURNED	30-Sep-2019

GENERAL COMMENTS	The paper is well written, clear, concise. To my opinion there is only one major issue: the comparison of stage distribution is unclear and needs better explanation. Sometimes it refers to a within-cohort comparison (i.e. looking at CPES and CR stage distributions separately), sometimes it refers to between-cohort comparison (i.e. stage by stage). The ORs compare the odds (in CPES cohort) of being diagnosed with a given stage compared to stage I and results vary according to the cancer type: Breast and colorectal patients are more likely to be diagnosed at stage II and III and less likely to be diagnosed at stage IV; lung patients are more likely to be diagnosed at stage II and less at stage III and IV; prostate patients are more likely to be diagnosed at any stage than stage I. On the other hand, if we compare the stage distribution in the CPES cohort with the stage distribution in the CR cohort we observe the following: a) Missing stage is more frequent in the CR cohort, and it should be interesting to elaborate on this b) Comparison between cohorts is very variable Hence, the conclusion “Survey responders are diagnosed with earlier stage disease” is questionable Furthermore, there are some minor comments which need to be addressed: Fig 1: excluded records from CR do not sum up to the correct number Fig. 2: not clear the purpose of this plot. What does “all registration” mean?

	p.6 line 10: add a sentence here to provide the total number of participants to the study. p.7 lines 1-2: not clear: did you impute missing stage? Was stage missing dealt with only in lung cancer? p.17 lines 19-21: for breast and prostate cancers, not only Asians have a lower likelihood than Whites to be in the CPES cohort, but also Blacks (prostate and breast) and Chinese (breast only) p.17 lines 22-24: your argument on stage differences is ambiguous: you should clarify that comparison is between the two cohorts and not within the same cohort. In any case your statement (CPES cohort more likely to have stage II and II disease and less likely to have stage IV disease) holds for colorectal and breast cancers only (looking at the 95%CI of ORs). p.17 lines 22-24: I find interesting the difference in unknown stage between the two cohorts. Any idea why? p.18 line15: Survey responders are not more likely to be diagnosed at stage II and III (see general comment). p.19 line 8-10: the stage distribution in the CPES cohort changes according to the cancer type and a unifying comment is not appropriate: see general comment p. 19 lines 13-19: not clear: where do we see the proportions of a cancer type wrt all types?

REVIEWER	Yoryos Lyratzopoulos UCL I have no financial conflict. I know some of the authors, particularly AD and ML, who are professional colleagues active in cancer healthcare epidemiology research, we have met many times in meetings and conferences, although we have not worked closely together in projects or in research income generation / holding grants together. I have published 16 papers using CPES data.
REVIEW RETURNED	20-Oct-2019

GENERAL COMMENTS	Thanks for asking me to review study, which comes from professional colleagues of mine with expertise in cancer patient experience research. The world has fallen in love with cancer patient experience surveys and cancer patient-reported outcome measure (cancer PROM) surveys, which is very positive. But some of the methodological issues such as those covered in the paper have not been fully realised. It is therefore important to generate the necessary evidence and the team (including patient representatives) should be congratulated for the work. Very good characterisation of the sample with regard to years of diagnosis covering 95% of the respondents and how those differ by cancer. The case-control approach is also excellent. I have few minor points for reflection by the authors. 1. Strengths and limitations of the study – key points '….wider cancer community diagnosed with the most common cancers in England'. And there is also a similar occurrence of this phrasing in-text I think. Is the term community best replaced by population of cancer patients or similar. 2. I was intrigued by the exclusion of patients diagnosed as death-certificate-only –is this applicable only to the 'controls', as CPES-sampled patients should not be DCO-diagnosed, or otherwise

	could the authors explain/comment – the way it is written it is not clear. 3. Ethnicity data – 'self-reported in HES'. In CPES datasets there is both a HES-recorded ethnicity item and a self-reported ethnicity item (as part of the questions that the patient completed), so there is a choice to be made, and it seems that the authors chose the HES-recorded item - which is acceptable. I believe HES-recorded ethnicity is meant to be self-reported but it is seems not to be always the case, as we have published a paper that shows that agreement between HES-recorded ethnicity and self-reported ethnicity within CPES (as part of the questionnaire item) is poor at the level of 16-group classification, though OK at major ethnic group level as in this study https://www.ncbi.nlm.nih.gov/pubmed/23811171 . Ethnicity reported as part of the CPES survey should be considered to be of superior accuracy to that contained in HES, though it is less complete. Some reflection/ animation along those lines with reference the prior evidence I think is appropriate in Discussion without any reanalyses needed given this is not the main issue/exposure analysed, and the concordance between the two sources is high enough anyhow. 4. Tables 1-4, column heading 'Cancer Registry' – can the label be more exact that these are cases outside the CPES sampling period (whatever the exact label / statement will be) as everyone is in the Cancer Registry and otherwise this column is not sufficiently distinguishable from the CPES column. 5. The Discussion is all-good, but looked to me to be a bit long – perhaps a bit of editing down in certain sections is possible. Regarding the prior literature, this paper from Denmark also shows similar findings with regard to a patient experience survey (of lung cancer patients) https://www.ncbi.nlm.nih.gov/pubmed/30642456 - as methodologically speaking PROM surveys are also surveys of cancer survivors they are likely subject to the same selection biases. Therefore, I would urge the authors to consider including reference to this Danish paper, and also, in the implications section, to expand on the relevance of the findings for Cancer PROM surveys too, particularly as there is ongoing work in this area currently. All-in-all a very topical/useful and methodologically sound paper. None of my comments require re-analysis, and I hope it to see it published soon.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 comment 1:

The paper is well written, clear, concise. To my opinion there is only one major issue: the comparison of stage distribution is unclear and needs better explanation. Sometimes it refers to a within-cohort comparison (i.e. looking at CPES and CR stage distributions separately), sometimes it refers to between-cohort comparison (i.e. stage by stage). The ORs compare the odds (in CPES cohort) of being diagnosed with a given stage compared to stage I and results vary according to the cancer type: Breast and colorectal patients are more likely to be diagnosed at stage II and III and less likely to be diagnosed at stage IV; lung patients are more likely to be diagnosed at stage II and less at stage III and IV; prostate patients are more likely to be diagnosed at any stage than stage I. On the other hand, if we compare the stage distribution in the CPES cohort with the stage distribution in the CR cohort we observe the following: a) Missing stage is more frequent in the CR cohort, and it should be

interesting to elaboration this b) Comparison between cohorts is very variable. Hence, the conclusion "Survey responders are diagnosed with earlier stage disease" is questionable.

Response comment 1:

Thank you for raising these important points. The results section has now been updated to clarify the comparison - between cohorts - and to include these findings. We agree that stage at diagnosis varies between cancers, but we think this is to be expected given the diverse cancer types included. This variation in the missing stage proportion is assumed to be due to two factors and we included them in the discussion: severity of the disease (poor prognosis patients are being missed by CPES), and the fact that the CPES responders are more users of the health system and therefore their data is well recorded. We therefore still believe that the survey is more likely to include patients with earlier stage disease although this was clearer among some cancer types than others.

Reviewer: 1 Comment 2:

Furthermore, there are some minor comments which need to be addressed:

Fig 1: excluded records from CR do not sum up to the correct number

Fig. 2: not clear the purpose of this plot. What does "all registration" mean?

p.6 line 10: add a sentence here to provide the total number of participants to the study.

Response to comment 2:

Thank you for raising these issues in our manuscript. Graphs shape, numbers, and legends were updated to clarify the work done in this manuscript. The text has been revised as suggested to the total number of participants to the study.

Reviewer: 1 Comment 3:

p.7 lines 1-2: not clear: did you impute missing stage? Was stage missing dealt with only in lung

Response to comment 3:

Stage was not imputed; only missing stage information was extracted from other sources within the cancer registry where it is available for lung cancer. The text has been revised to reflect this.

Reviewer: 1 Comment 4:

p.17 lines 19-21: for breast and prostate cancers, not only Asians have a lower likelihood than

Whites to be in the CPES cohort, but also Blacks (prostate and breast) and Chinese (breast only)

Response to comment 4:

Thank you for pointing this. The results section was revised to reflect this association.

Reviewer: 1 Comment 5:

p.17 lines 22-24: your argument on stage differences is ambiguous: you should clarify that

comparison is between the two cohorts and not within the same cohort. In any case your

statement (CPES cohort more likely to have stage II and II disease and less likely to have stage IV disease) holds for colorectal and breast cancers only (looking at the 95%CI of ORs).

p.17 lines 22-24: I find interesting the difference in unknown stage between the two cohorts. Any idea why?

p.18 line15: Survey responders are not more likely to be diagnosed at stage II and III (see

general comment).

p.19 line 8-10: the stage distribution in the CPES cohort changes according to the cancer type

and a unifying comment is not appropriate: see general comment

p. 19 lines 13-19: not clear: where do we see the proportions of a cancer type wrt all types?

Response to comment 5:

Thank you for raising these points. We hope we have addressed all these comments in our response to the general comments.

Reviewer: 2 comment 1

Thanks for asking me to review study, which comes from professional colleagues of mine with expertise in cancer patient experience research. The world has fallen in love with cancer patient experience surveys and cancer patient-reported outcome measure (cancer PROM) surveys, which is very positive. But some of the methodological issues such as those covered in the paper have not been fully realised. It is therefore important to generate the necessary evidence and the team (including patient representatives) should be congratulated for the work. Very good characterisation of the sample with regard to years of diagnosis covering 95% of the respondents and how those differ by cancer. The case-control approach is also excellent. I have few minor points for reflection by the authors.

Response to comment 1:

Thank you.

Reviewer: 2 comment 2

1. Strengths and limitations of the study – key points '…. wider cancer community diagnosed

with the most common cancers in England'. And there is also a similar occurrence of this phrasing in-text, I think. Is the term community best replaced by population of cancer patients or similar?

Response to comment 2

Thank you for raising this point. “community” has now been changed to “population”.

Reviewer: 2 comment 3

2. I was intrigued by the exclusion of patients diagnosed as death-certificate-only –is this applicable only to the ‘controls’, as CPES-sampled patients should not be DCO-diagnosed, or otherwise could the authors explain/comment – the way it is written it is not clear.

Response to comment 3

Thank you. We agree that in reality CPES responders could not be DCOs. The text has been revised in the method to clarify this point.

Reviewer: 2 Comment 4

3. Ethnicity data – ‘self-reported in HES’. In CPES datasets there is both a HES-recorded ethnicity item and a self-reported ethnicity item (as part of the questions that the patient completed), so there is a choice to be made, and it seems that the authors chose the HES recorded item - which is acceptable. I believe HES-recorded ethnicity is meant to be self-reported but it is seems not to be always the case, as we have published a paper that shows that agreement between HES-recorded ethnicity and self-reported ethnicity within CPES (as part of the questionnaire item) is poor at the level of 16-group classification, though OK at major ethnic group level as in this study https://www.ncbi.nlm.nih.gov/pubmed/23811171 . Ethnicity reported as part of the CPES survey should be considered to be of superior accuracy to that contained in HES, though it is less complete. Some reflection/ animation along those lines with reference the prior evidence I think is appropriate in Discussion without any reanalyses needed given this is not the main issue/exposure analysed, and the concordance between the two sources is high enough anyhow.

Response:

Thank you for raising this important point. While we agree on that self-reported ethnicity varies between datasets and at different times, the decision on using ethnicity from HES was because we wanted a fair comparison between the two cohorts using the same source of ethnicity data.

Reviewer: 2 Comment 5

4. Tables 1-4, column heading ‘Cancer Registry’ – can the label be more exact that these are cases outside the CPES sampling period (whatever the exact label / statement will be) as everyone is in the Cancer Registry and otherwise this column is not sufficiently distinguishable from the CPES column.

Response to comment 5

Thank you for raising this important point. The text has been revised in the table label and legends to clarify this point.

Reviewer: 2 Comment 6

5. The Discussion is all-good but looked to me to be a bit long – perhaps a bit of editing down in certain sections is possible. Regarding the prior literature, this paper from Denmark also shows similar findings with regard to a patient experience survey (of lung cancer patients) https://www.ncbi.nlm.nih.gov/pubmed/30642456 -as methodologically speaking PROM surveys are also surveys of cancer survivors, they are likely subject to the same selection biases. Therefore, I would urge the authors to consider including reference to this Danish paper, and also, in the implications section, to expand on the relevance of the findings for Cancer PROM surveys too, particularly as there is ongoing work in this area currently.

Response to comment 6

Thank you for the comment and suggesting the recent paper, which has been now added into the reference list and mentioned in the study implications. Regarding the discussion formatting, two patient representatives (MB and JR), who themselves had conducted work on the CPES, were involved in the study methods and the manuscript writing. We asked again for their opinion on the discussion, and they feel that the slightly longer discussion reads well and is necessary for the wider audience of policy makers and user representative using CPES findings.

VERSION 2 – REVIEW

REVIEWER	Anna Gigli Institute for Research on Population and Social Policies, National Research Council - Italy
REVIEW RETURNED	08-Nov-2019

GENERAL COMMENTS	Generally, I believe that providing single answers to single issues should be preferred. Grouping the answers, as the authors did, makes it more difficult to check the changes in the text. Nevertheless, the revised version is quite satisfactory. There are a few mistakes in the text: 1. p. 6 line 10 in comparison with fig. 1: records excluded are 3673, and not all of them are excluded because they are multiple records. Please change the text 2. p. 8 line 28 in comparison with table 3: adjusted OR is 0.81. Please change the text 3. p. 17 line 30 in comparison with tables 1 and 3: ORs and CIs are inverted. Please change the text 4. The note at the bottom of table 1-4 “Cancer registry: This group of population does not include any CPES responders”: this fact is apparent in fig. 1 and I don’t think it necessary to repeat it here. However, if the authors wish to do so, they must provide a symbol (other than * that has already been used) and use the symbol in the table header, as well.

REVIEWER	Yoryos Lyratzopoulos UCL and PHE
REVIEW RETURNED	06-Nov-2019

GENERAL COMMENTS	The authors addressed the comments and improved the manuscript.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1 comment 1:

p. 6 line 10 in comparison with fig. 1: records excluded are 3673, and not all of them are excluded because they are multiple records. Please change the text.

Response to comment 1:

Thank you for raising this point. We have updated the text.

Reviewer: 1 comment 2:

p. 8 line 28 in comparison with table 3: adjusted OR is 0.81. Please change the text

Response to comment 2:

Thank you for spotting this typo. We have corrected this minor mistake.

Reviewer: 1 comment 3:

3. p. 17 line 30 in comparison with tables 1 and 3: ORs and CIs are inverted. Please change the text

Response to comment 3:

Thank you for raising this point to us. We have corrected the mistake and updated the text.

Reviewer: 1 comment 4:

The note at the bottom of table 1-4 “Cancer registry: This group of population does not include any CPES responders”: this fact is apparent in fig. 1 and I don’t think it necessary to repeat it here. However, if the authors wish to do so, they must provide a symbol (other than * that has already been used) and use the symbol in the table header, as well.

Response to comment 4:

Thank you for raising this point. We agree, and therefore deleted this note from all tables.